



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

The Emergence of Institutions

Citation for published version:

Sanchez-Pages, S & Straub, S 2010, 'The Emergence of Institutions', *The B.E. Journal of Economic Analysis & Policy*, vol. 10, no. 1, 84. <https://doi.org/10.2202/1935-1682.2465>

Digital Object Identifier (DOI):

[10.2202/1935-1682.2465](https://doi.org/10.2202/1935-1682.2465)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Publisher's PDF, also known as Version of record

Published In:

The B.E. Journal of Economic Analysis & Policy

Publisher Rights Statement:

With permission. © Walter de Gruyter GmbH. Sanchez-Pages, S., & Straub, S. (2010). The Emergence of Institutions. *The B.E. Journal of Economic Analysis & Policy*, 10(1), [84]doi: <http://dx.doi.org/10.2202/1935-1682.2465>

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



The B.E. Journal of Economic Analysis & Policy

Contributions

Volume 10, Issue 1

2010

Article 84

The Emergence of Institutions

Santiago Sanchez-Pages*

Stéphane Straub[†]

*University of Edinburgh, ssanchez@staffmail.ed.ac.uk

[†]Université de Toulouse, stephane.straub@univ-tlse1.fr

Recommended Citation

Santiago Sanchez-Pages and Stéphane Straub (2010) "The Emergence of Institutions," *The B.E. Journal of Economic Analysis & Policy*: Vol. 10: Iss. 1 (Contributions), Article 84.

The Emergence of Institutions*

Santiago Sanchez-Pages and Stéphane Straub

Abstract

This paper analyses how institutions aimed at coordinating economic interactions may emerge. Starting from a hypothetical state of nature, agents can delegate the task of enforcing cooperation to one of them in exchange for a proper compensation. Both individual and collective commitment problems stand in the way of institution formation. These problems imply first that a potentially efficient institution may fail to emerge and also that if it emerges, it may do so inefficiently. We show that big and untrustworthy societies are more likely to support institutions whereas their emergence is more difficult in small and trusting societies, but if institutions do emerge, they tend to be more inefficient in the former type of societies. Finally, we show that the threat of secession by a subset of agents may alleviate the latter problem.

KEYWORDS: institution, coordination, state of nature, secession

*We thank the editor Antonio Cabrales as well as three anonymous referees, Andrzej Baniak, Eric Brousseau, Paco Candel-Sánchez, Avinash Dixit, Jan Fidrmuc, Francesco Giovannoni, Chris Kingston, Richard Langlois, Patrick Legros, John Moore, Boaz Moselle, Luis Garicano, Peter Grajzl, Emmanuel Raynaud, Gilles Saint-Paul, József Sákovics, Paul Seabright, Jonathan Thomas, Jean Tirole, Charles Vellutini, Irenaeus Wolf and seminar audiences at Edinburgh, Birmingham, Toulouse, CEU Budapest, Esnie 2006, ISNIE 2007, and the 11th Coalition Theory Workshop in Warwick for their useful comments. Sánchez-Pagés also acknowledges financial support from the Spanish Ministry of Education and Science, grant SEJ2005-02079.

“As for ‘philosophical history’, it involved accounting for the development of beliefs, practices, theories, and institutions on the basis of natural causes or principles, when actual records and reports of witnesses were lacking.”

Ian Simpson Ross, *The Life of Adam Smith* (1995).

1 Introduction

1.1 Motivation and overview

To date the literature on the economic analysis of social and political institutions has focused mainly on their role as protectors of property rights.¹ A more neglected role of institutions is to correct the coordination failures or commitment problems that sometimes plague the most basic type of economic interactions. Contributions along this line of enquiry have modeled institutions as a set of self-enforcing constraints on behavior, usually relating them to narrative descriptions derived from historical sources.²

These works describe institutional arrangements already in place, but very little has been said on the factors that lead to the emergence of these institutions. The aim of this paper is to analyze the process of creation of a specialized institution responsible for enforcing efficient rules for social interactions. We do this in the context of “individualist” societies, as Dixit (2003a) puts it, in which interactions mostly involve strangers and cooperation cannot rely on the type of multilateral mechanisms, mostly norms and communication, that prevail in more “collectivist” societies such as families, or small and homogenous ethnic or religious groups. Our model of institution formation especially applies to the emergence of institutional intermediaries in trade exchanges in early societies and intra-industry interactions between firms. Examples of the former include the introduction of commercial witnesses in the development of formal trade between tribes in Africa, aboriginal Australia or pre-Columbian Nicaragua (Attali, 2003).³ Examples of the latter include the emergence of the private legal framework that rules the US cotton industry (Bernstein, 2001) or the private arrangements in the diamond industry (Bernstein, 1992, Richman, 2006), the creation of formal credit reference agencies in Ghana, Kenya and Zimbabwe (Fafchamps, 2004), and of dispute resolution mechanisms in Asia (MacMillan and Woodruff, 1999, 2000).

¹See Bardhan (2005).

²See Greif (1997) for a review.

³See also Milgrom, North and Weingast (1990), Dixit (2003a).

The point of departure of our model is an economy in the state of nature in which the value of each individual's endowment is enhanced through bilateral interactions. Interactions take the form of a simple prisoners' dilemma game. Mutual cooperation would benefit both parties but being opportunistic is a dominant strategy and in equilibrium low payoffs are realized. In this context, the institution is a body capable of enforcing the Pareto efficient outcome in any bilateral interaction. For the sake of tractability, and because we want to focus on the potential emergence of this body, we choose not to model explicitly its internal functioning.

The process of emergence of such institution is modelled through a dynamic, non-repeated game that agents play in this hypothetical state of nature. Depending on the existing incentives, players' actions will eventually lead to the establishment of such formal coordinating mechanism as an equilibrium of the game. For the formal institution to arise, agents need to delegate to one of them the task of running it. The institution is costly to set up. The delegate must relinquish her ability to interact with other agents, and must be properly compensated in exchange. If the institution arises, agents have to decide whether to become formal and abide by its norms of interaction or not. Whenever two formal agents meet, the institution can guarantee that the efficient outcome will result. However, in order to enjoy this benefit, agents must pay a fee that constitutes the source of revenue for the institution.

We explore several procedures of institution formation and characterize under which circumstances they will be successful. We make special emphasis on the impact of these different processes on efficiency and welfare.

Individuals' motivation to participate in the process of institution formation are the rents associated with the institutional task. This, together with the fact that they do not fully internalize the social benefit generated by the institution, implies that a decentralized process of institution formation is plagued by two commitment problems. The first one is simply the individual commitment problem that arises when the revenue that can be raised by the agent chosen to act run the institution is insufficient, and she prefers to renege ex post, leading society to fall back into informality. The second one, which we label "collective commitment" problem, is linked to the fact that agents may not be able to write binding agreements on the fee that will be charged by the institution once it is created.

Both limitations on commitment have implications for efficiency. The first one implies that an institution may not arise due to the lack of individual incentives and despite being potentially welfare enhancing. This is in particular the case when the extent of the coordination problem is limited. The intuition is that when the level of trust in the state of nature is relatively

high, the outside option in which no institution emerges is more attractive and agents' willingness to pay in order to create the institution is lower. On the other hand, the lack of collective commitment implies that even if an institution emerges, it may do so at a sub-optimal level of efficiency because the fee finally charged may be too high from a social point of view. This happens for low levels of trust in the state of nature, because in that case the institution is able to set a high fee compared to the first-best level.

We show that small societies with high levels of trust are less likely to support the emergence of institutions than big ones with low levels of trust, but if institutions do emerge, they are more likely to be inefficient in the latter type of societies.

Exogenously imposed commitment along each one of the two dimensions alone would reduce the scope for inefficiencies, but the first-best institution emerges only when both problems can be solved simultaneously. We then examine several devices that may help to solve these commitment problems endogenously. The first one is agents' use of a particular type of strategies that enable them to constrain the ability of the institution to extract too much resources. The second potential improvement is the threat of secession by a subset of agents. To deter secession, the institution must charge a fee that cannot be improved upon by any coalition. However, this effect only operates for a limited parameter space; a big population size and high levels of trust in the state of nature make it very attractive to become a central agent and therefore create too strong incentives to secede.

The remainder of the paper is as follows. Next, we review the related literature and spell out our specific contribution in that context. Section 2 presents the model and its basic elements. In Section 3 we characterize the equilibrium level of formality, once the institution has arisen, and the first best fee from the viewpoint of a social planner. Section 4 explores different procedures of institution formation characterized by varying degrees of commitment. In Section 5 we explore two ways to endogenize commitment. Section 6 discusses the results and concludes. Proofs are relegated to the Appendix.

1.2 Literature review

Previous works on the role of institutions as coordination devices have mainly explored the functioning of specific institutional arrangements, in the light of both empirical accounts and game-theoretical modelling.⁴ In the words

⁴See Greif (1997) for a survey of the economic history literature that relies on micro-economic theory to study institutions.

of Dixit (2004), in this literature a distinction can be drawn depending on how the enforcement of coordination or cooperation is attained: On the one hand, models of self-governance, understood as equilibria sustained by some type of relation-based multilateral mechanism (punishment, communication), and on the other hand models of formal rule-based institutional arrangements, whereby some qualified agent(s) takes on the role of solving coordination problems among the rest of the population. The former is exemplified by Greif's (1993) study of the coalition supporting the interactions of Maghribi traders with their distant agents in the 11th century, and modelled for example in Kandori (1992) and Ellison (1994), while case studies of the latter, characterized by bilateral enforcement of interactions, are found in Milgrom et al. (1990), Bernstein (2001), and Richman (2006) among others. Relation-based multilateral enforcement generally prevails in relatively homogenous groups, while bilateral rule-based enforcement mechanisms are more likely in the context of relatively anonymous interactions in large groups.⁵

Our analysis focuses on formal rule-based institutions. Apart from the ones mentioned above, examples include the case of organized crime and the role of Peppe in ensuring that cattle breeders and butchers do not cheat each other when trading (Gambetta, 1993) or the merchant courts at the Champagne fairs of the 12th and 13th centuries (Milgrom et al., 1990).⁶ Such institutions are potentially very diverse and can rely not only on coercion, as in the "Hobbesian" approach of Bates, Greif and Singh (2002), but also on more subtle forms of persuasion, e.g., as a "Humean" *political entrepreneur* or government, able to persuade others to take a particular action or alter their beliefs about this action's consequences (see Taylor, 1982, and Basu, 2000).

How do these institutions emerge? The literature on the origins of institutions distinguishes two main views. On the one hand is the "efficient institutions" view, whereby optimal institutions emerge as they are in the interest of the whole community. Such efficient institutions, addressing different underlying failures or problems, are found for example in Grossman (2002), Olson's (2000) roving versus stationary bandits and Baker, Bulte and Weisdorf (2010) and Baker (2003). These papers address situations in which some agent is endowed with power in order to protect the rest of the community against an outside threat. Bates et al. (2002) provide numerous historical examples to support their theoretical account of why agents may seek to attribute the monopoly of violence to a delegate, a canonical State, in charge of ensuring

⁵Greif (1994). Dixit (2003b) and Li (2003) provide theoretical foundations.

⁶See also Greif (1997) on Genoese traders, and Konrad and Leininger (2007) on clan leaders in inter-group conflict.

peace and enabling higher levels of production. More generally, accounts of the State as an entity solving coordination problems also fit here.⁷ On the other hand, the “social conflict perspective” holds that potentially inefficient institutions emerge and evolve as the result of a struggle between different interest groups, as modelled for example in Acemoglu and Robinson (2005).⁸

Our contribution falls in the “efficient institutions” category of models. However, it displays a number of differences with the existing literature. First of all, while in most papers mentioned above, the inefficiency rooted in the decentralized setting is the exposure to violence or robbery, leading to the emergence of an institution that taxes people in exchange for protection, we focus on a situation in which the fundamental inefficiency is the lack of incentives to cooperate. In that sense, our paper formalizes some of the claims made informally by Greif (2005) who, based on historical examples, concludes that societies with strong informal institutions will have a low demand for formal enforcement methods. The author also shows how rulers’ ability to abuse rights hinder the creation of institutions. In our case, the ruler’s ability to extract the whole surplus has an impact on efficiency rather than on emergence, leading to a situation in which institutions may still emerge in low trust societies but they are more likely to be inefficient. Furthermore, this extracting ability is the key reason why some agent will be willing to perform the institutional task, although it may not always be sufficient.

Second, within that class of models, we analyze departure from the first best due to problems of commitment. In that sense, we are able to characterize not only conditions under which an institution emerges to address a fundamental inefficiency, but also how limited commitment, both on the side of the chosen leader and of the general population, leads to second-best outcomes. This means that institutions may be too extractive, but also that they may not be immune to the risk of secession and could therefore end up benefiting only subgroups of the population. Indeed, in our model the institution initially emerges from the set of agents that would like to run it, but we later consider the possibility that groups may secede from society in order to create their own institution if they find that the original one was too extractive. In that sense, our conclusions can for example be related to the

⁷See Taylor (1982) and Basu (2000). This is also an important topic in social anthropology. For example, Ensminger (1992) describes the century-long process through which changes in the environment triggered the Orma tribe in Kenya to move from a rule by a council of elders to the recognition of the authority of the modern Kenyan nation-state.

⁸Some contributions have analyzed how informal or personalized relationship-based institutions may coexist with more formal, rule-based anonymous mechanisms, and how the transition from one to the other may occur (e.g., Kranton, 1996, and Dixit, 2004).

account in Fafchamps (2004), who shows that although in Zimbabwe, contrary to Ghana and Kenya, the coordination problem between firms mentioned above had apparently been solved through the creation of a formal credit reference agency, the outcome was not as efficient as expected. Indeed, the institutional mechanism in fact only benefited a limited number of firms, defined along ethnic boundaries, and was de facto excluding some of the players who stood to benefit from it.

Third, our methodological perspective is new in the sense that, in contrast to the existing literature, our model explicitly focuses on the emergence process, by endogenizing the rise of a ruler from a population of identical individuals, in contrast with other works in the literature that exogenously impose its existence (Acemoglu, 2003; Acemoglu et al., 2004) and look at how its presence shapes economic outcomes, or compare the scenarios with and without ruler (Grossman, 2002; Moselle and Polak, 2001; Baker, 2010) to characterize parameters thresholds associated with potential emergence.

Finally, we also contribute to the discussion on the link between social capital and institutions (Durlauf and Fafchamps, 2005, and Ahlerup et al., 2009), by providing micro-foundations for the observed negative correlation between trust and the likelihood of formal institution emergence.

2 The model

Consider an economy populated by $N + 1$ agents, each having an initial endowment ω (representing a combination of skills, time and goods). Agents' interactions in this economy are described by the basic game G in Figure 1.

		Player j	
		C	NC
Player i	C	x, x	$-z, z$
	NC	$z, -z$	$0, 0$

Figure 1: Basic game G .

Agents are anonymous to each other. They are randomly and bilaterally matched to play G . Payoffs in the matrix represent the return per unit of

endowment invested in the interaction. We assume that $z > x > 0$. The strategy C stands for a cooperative behavior that can create added value, and NC stands for an opportunistic behavior that allows the agent to take advantage of a cooperating partner but yields zero returns otherwise.

The game G admits a unique Nash equilibrium, (NC, NC) , that is Pareto inferior to (C, C) . As is common in the literature, we have chosen this game as the simplest way to illustrate the type of coordination problems that occur in many social situations, but many other games could capture the same trade-offs we study here, like for instance a coordination game with two equilibria, one Pareto-superior to the other.

The scenario in which individuals are randomly and bilaterally matched and play G without any interference is assumed to be the status-quo of the economy. In order to solve the problem of opportunism, agents can set up an institution with the power to enforce cooperation and ensure that the efficient outcome (C, C) results from any interaction between agents under its oversight. Given that we are mainly interested in studying when such institutions can arise, and that in doing so we want to keep the analysis tractable, in what follows we just consider a reduced form of the actual process the institution employs to enforce cooperation.

This institution arises when agents delegate to one of them, who we will call the *center*, the task of running it. The central agent relinquishes her ability to interact with other agents, so she must be compensated in exchange. At this point, we deliberately remain vague about how this delegation process is carried through since the main body of the paper (Section 4 below) amounts to discussing several procedures of institution formation.

If the institution arises, agents have to decide whether to abide by it, that is, to become formal, or not to do so and remain informal. However, in order to become formal and interact under the institutional umbrella, they each have to pay a fixed fee $a \leq \omega$, that can be understood as an entry fee or a lump-sum tax that rewards the center for her activity. Below we will also discuss at length how the level of the fee a is fixed.

We will admit a richer description of the payoff x in G and assume that it depends on the efficiency of the institutional mechanism that in turn is a function of the level of agents' individual contribution a .⁹ Hence, the per-person unit return from an interaction between two formal agents is

⁹Making x also dependent on N would introduce scale effects. Alternatively, x could also be a function of the proportion of agents $\frac{K}{N}$ contributing to it. However, it is unclear how this would affect x . Indeed, a higher proportion could have a positive effect because of network externalities for example, but congestion could also lead to a negative effect (see Kranton, 1996).

$$v^F = x(a), \quad (1)$$

where the superscript F denotes “Formal” and $x(\cdot)$ satisfies $x_a > 0$, $x_{aa} < 0$, and the standard Inada condition, $\lim_{a \rightarrow 0} x_a(a) = \infty$. We thus assume that the institution becomes more efficient when endowed with more resources, as it is able to monitor better its members’ behavior or to invest more in physical or relational supporting infrastructure, as for example in the case of diamond clubs described in Richman (2006). If x did not depend on a , so for instance it took two values depending on whether the institution arises or not, all our results on emergence would hold, but we would be unable to obtain different levels of efficiency.

When at least one of the two interacting agents is informal, the institution has no power to enforce the efficient outcome and the game G is played without any further interference. Informal agents thus avoid paying the fee but their interactions yield lower returns. More formally, the per-person *expected* unit return when G is played between a formal and an informal agent or between two informal agents is

$$v^I = y, \quad (2)$$

where the superscript I denotes “Informal”. This payoff can be interpreted as the upper bound on what can be achieved between agents who do not use the formal institutional umbrella. For example, when there exist unforeseen fluctuations in the environment, these fluctuations may create “temptations” to act non-cooperatively.¹⁰ In the absence of an external institution capable of solving this problem, such temptations are optimally addressed by settling for lower cooperation payoffs. Alternatively, this payoff can be interpreted as the payoff of self-enforcing agreements under finitely repeated interactions when players believe that a fraction of agents are altruists or use tit-for-tat strategies.¹¹

To keep the model tractable, y is assumed to be an initial condition of our economy. More specifically, we define $\tau \equiv \frac{y}{x(a)} < 1$, as a measure of the level of cooperation attainable in a world without institution relative to the situation where the institution does arise. Under our first interpretation, this is equivalent to assume that the informal payoff y is a constant proportion of the one obtained under the institution for any given fee $(x(a))$. Under the second interpretation, the measure τ can be related to the share of the population

¹⁰In Rotemberg and Saloner (1986) these fluctuations emerge from changes in demand that create incentives to renege on collusive agreements. Similarly, Bagwell and Staiger (1990) study international trade when trade swings create incentives for protectionism.

¹¹See Kreps et al. (1982).

believed to have altruistic preferences. Kreps et al. (1982) show that the extent of cooperation is non-decreasing in this share. This degree of cooperation is also often referred to as a measure of “social capital”.¹² Again, our assumption in this case would be that the degree of cooperation does not change with the fee (i.e., that y is a constant fraction of $x(a)$). A third interpretation of the parameter τ , in line with the literature on informality, is the level of free-riding that informal interactions can make on formal institutions.¹³ In what follows, we will refer to τ as the level of trust or cooperation in the society under the state of nature.

Finally, we assume that $x(0) > \frac{1}{\tau}$ to ensure that participating in a completely informal economy always dominates the autarchic situation in which agents do not interact and simply consume their endowments.

Under risk neutrality and anonymous random matching,¹⁴ the expected payoff of a formal agent when $K \geq 2$ agents are formal can be written as

$$\begin{aligned} V^F &= \frac{K-1}{N-1} (\omega - a) v^F + \frac{N-K}{N-1} (\omega - a) v^I \\ &= \frac{K-1}{N-1} (\omega - a) x(a) + \frac{N-K}{N-1} (\omega - a) \tau x(a). \end{aligned} \quad (3)$$

We assume that an institution becomes active if at least two agents are formal, so the probability of formal exchanges is strictly positive. On the other hand, the payoff for an informal agent is just.

$$V^I = \omega v^I = \omega y = \omega \tau x(a),$$

since he avoids paying the formality fee. Note that when no more than one agent becomes formal, $v_a^F = v_a^I = \tau x(0)$ for any a that the central agent might have set.

Finally, the central agent, who gives up interacting with the rest of agents, receives the fees paid by all formal agents. Her payoff is thus given by

$$V^C = K(a - c),$$

where c is the fixed enforcement cost per formal agent, linked for example to the need to record information on its behavior, maintain proper infrastructure, etc. As will come clear below, an alternative formulation with a

¹²Routledge and von Amsberg (2003), Durlauf and Fafchamps (2004).

¹³See for example Marcouiller and Young (1995), Choi and Thum (2002), Azuma and Grossman (2002) and Straub (2005).

¹⁴We discuss alternative matching technologies in the final section of the paper.

cost proportional to the number of formal transactions would not qualitatively change results.

Figure 2 summarizes the timing of the game described above.

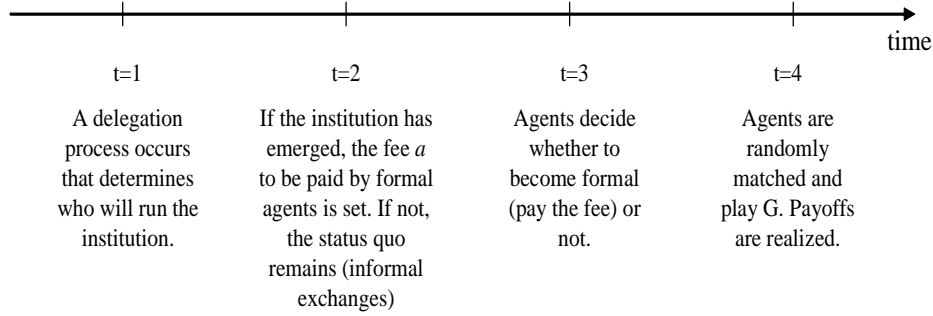


Figure 2: Timing of the game.

We have thus constructed a dynamic, non-repeated game of complete information in four stages. In the first stage, agents set up the institution (the game stops there if they do not succeed in doing so). Then, the institutional fee a is set. In the third stage, agents decide whether to become formal or not. In the last stage, they are paired with another interacting agent in society and play G , eventually resorting to the institution set up earlier.

3 The equilibrium level of formality

3.1 Existence and stability

Given this basic framework, the first question that arises concerns the existence and stability of different configurations. Assume that K agents are formal, $N - K$ are informal, and that, without loss of generality, the $N + 1^{th}$ agent is devoted to institutional work. Given a fee a , this division of agents between formality and informality can be supported in equilibrium if and only if no agent is willing to deviate and change her status.

A formal agent's best response is not to deviate and become informal as long as $V_{K,a}^F \geq V_{K-1,a}^I$. After some transformations, this can be written:

$$a \leq \omega \left(1 - \frac{(N-1)\tau}{K-1 + (N-K)\tau} \right) \equiv a(K).$$

Similarly, an informal agent will not choose formality as long as $V_{K,a}^I \geq V_{K+1,a}^F$, which yields:

$$a \geq \omega \left(1 - \frac{(N-1)\tau}{K + (N-K-1)\tau} \right) = a(K+1).$$

Note first that $0 < a(K) < \omega$ for all $K > 1$ and that given our assumption above stating that the institution remains inactive if $K = 1$, $a(1) = 0$.

The equilibrium level of formality will depend upon the properties of $a(\cdot)$. The next Proposition characterizes the conditions under which there exists a level of the institutional fee a that can support a certain amount of formal agents in equilibrium.

Proposition 1 *The function $a(\cdot)$ is strictly increasing in K . Hence, there are only two equilibria, one with $K = 0$ and one with $K = N$.*

In this model, only corner equilibria exist, i.e. full formality or full informality¹⁵. When full formality prevails, $a(N) = \omega(1 - \tau)$. We will assume that $c < \omega(1 - \tau)$. Otherwise even the highest fee compatible with full formality could not cover the running costs of the institution. The following proposition characterizes a necessary condition for this outcome to be an equilibrium.

Proposition 2 *A necessary condition for informality ($K = 0$) to be an equilibrium is $a \geq 0$, whereas a necessary condition for full formality ($K = N$) to be an equilibrium is $a \leq a(N)$.*

The proof follows from the arguments above. This Proposition shows that a coordination problem arises when the institution emerges. Paying a fee compatible with full formality may not compensate the cost of becoming formal when everybody else is informal. Hence, both full formality and informality can be sustained in equilibria for the same level of the fee. For the rest of the paper, we will mainly focus on the equilibrium in which the institution forms. We see it as more natural because at that point of the game agents have decided to participate in the process, a central agent has been chosen and the fee has already been announced. Still, we will leave a further discussion on the informal equilibrium to Section 5.

¹⁵In Kranton (1996), a similar outcome arises, in which the entire population engages either in market or in reciprocal exchange.

Figure 3 depicts the profile of equilibria as a function of the fee a .¹⁶

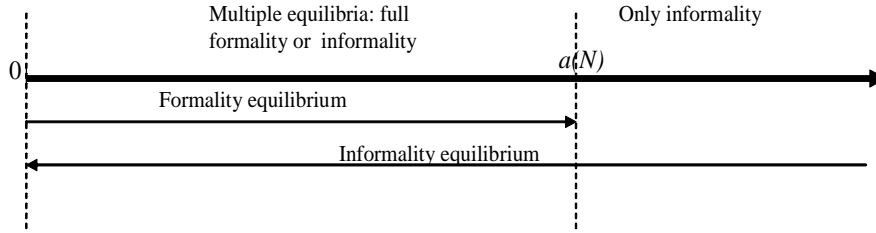


Figure 3: Profile of equilibria

3.2 The first best institutional fee

In the remainder of this Section, we characterize the optimum fee from the viewpoint of a hypothetical central planner willing to maximize the total sum of agents' utilities. The planner compares the maximum welfare attainable in the scenario in which agents pay a fee to enjoy the benefits of formal interactions with the (fixed) level of social welfare under complete informality.

In the case of full formality, the constrained maximization problem of this planner can be written as:

$$\begin{aligned} \max_a W^F(a) &= N[(\omega - a)x(a) + (a - c)] + \omega \\ \text{s.t.} \quad a &\leq a(N). \end{aligned}$$

Given our assumptions on $x(\cdot)$, this objective function is concave. The intuition behind this is that increasing a makes the institution more efficient but it also reduces the endowment available for interactions. The first order condition implicitly characterizes an interior solution a^* such that:

$$\omega - a^* = \frac{x(a^*) - 1}{x_a(a^*)}. \quad (4)$$

¹⁶For the range of fees $[0, a(N)]$ there also exists a mixed strategy equilibrium in which agents become formal with probability $p(a) = \frac{\tau}{1-\tau} \frac{a}{\omega-a}$. Although this can in principle support an intermediate level of formality, the revenue raised by the institution in this equilibrium is maximized at $a = a(N)$ so $p(a(N)) = 1$. Hence, in this case full formality would arise as well.

The fee a^* cannot be higher than the maximum fee compatible with full formality. Hence, the planner chooses to implement full formality with a fee equal to

$$a^F = \min\{a^*, a(N)\}.$$

This implies that the solution to the planner's problem will be a corner solution, i.e. $a^* \geq a(N)$, as long as $\tau \geq \tau^*$, where τ^* satisfies:¹⁷

$$\tau^* = \frac{x(\omega(1 - \tau^*)) - 1}{\omega x_a(\omega(1 - \tau^*))}. \quad (5)$$

As τ increases, formal agents have stronger incentives to defect. A lower fee $a(N)$ is now necessary to support formality and the room for an interior solution shrinks. On the other hand, the effect of an increase in the endowment ω is ambiguous: It relaxes the constraint but also changes the objective function by making interactions more profitable.

The planner can leave the economy in a state of full informality. In that case, total welfare is just

$$W^I = (N + 1)\omega\tau x(0). \quad (6)$$

There may exist values of the parameters for which the maximum welfare under formality, i.e. $W^F(a(N))(a^F)$, is lower than under informality. This is characterized by the following Proposition.

Proposition 3 *There exists a threshold \underline{x} on the payoff under autarky $x(0)$ such that social welfare under full formality is lower than under informality if*

$$x(0) > \underline{x} \equiv \frac{1}{N + 1} \left(\frac{1}{\tau} + N \frac{(\omega - a^F) x(a^F) + a^F - c}{\tau\omega} \right). \quad (7)$$

Moreover, \underline{x} is increasing in both the population size N and agents' initial endowment ω and decreasing in the status-quo level of trust τ .

Hence, for small and relatively poor economies (low N or ω) full formality may not be the most desirable outcome. Similarly, when the problems of miscoordination are not very severe (high τ), the cost of setting an institution may outweigh the gains.

¹⁷It is straightforward to show that such fixed point exists.

4 Emergence of the institution

Since no central coordination device exists before the members of a society actually create one, any effort to set up an institution that will enforce co-operation has to proceed in a decentralized way. The basic idea is that the center must be given incentives to act as such. In this Section, we analyze this process and highlight how commitment problems affect the efficiency of the emerging institution or block its emergence despite its potentially welfare enhancing effect.

We define a *procedure of institution formation* as a fair lottery over the set of agents who freely participate in it. This lottery designates the agent who subsequently will be in charge of running the institution. The fee a can be set either before the lottery takes place or afterwards.

A lottery is the simplest mechanism to choose the individual to be in charge of the institution. It is so because all agents are alike and hence there is no reason why *a priori* they should not be equally likely to end up in charge of the institution. This is reminiscent of Rawls' (1971) concept of the *original position*. Moreover, a lottery also constitutes the reduced form of more complex mechanisms, like for instance auctions or contests: If they were used, the equilibrium would be typically symmetric and a winner should therefore be randomly chosen.

In this general framework, different procedures of institution formation are possible depending on the different degrees of commitment available both at the individual and at the collective level, the natural benchmark being a fully decentralized process with no commitment whatsoever.

Agents must decide simultaneously and *ex-ante* whether to participate or not in the lottery that will designate who will run the institution. Hence, given a level of the fee a , the institution can arise only if

$$\frac{1}{N+1}(N(a-c) + \omega) + \frac{N}{N+1}(\omega - a)x(a) \geq \omega \tau x(a), \quad (8)$$

where the left hand side shows the expected payoff from participating, as the sum of the center's and the agents' payoffs respectively weighted by their corresponding probabilities, and the right hand side is the payoff from unilateral deviation. In equilibrium, it is easy to show that either all or no agent will participate in the process.

All the different processes of institution formation that we will discuss next rely on this basic participation constraint. Still, agents who accept to participate in the process may change their mind *ex-post* depending on the outcome of the lottery. Therefore, when there is no commitment at the individual level, an *ex-post* participation constraint needs to be imposed as well. This requires that, once they discover their role, agents should not prefer to fall back into informality. As this *ex-post* requirement is always satisfied for an agent who does not become the center¹⁸, the center's *ex-post* participation constraint is the only relevant one:

$$N(a - c) + \omega \geq \omega \tau x(0). \quad (9)$$

This defines a minimum level of the fee

$$\underline{a} \equiv \omega \frac{\tau x(0) - 1}{N} + c,$$

below which the agent chosen to be the center would prefer to give up and the whole economy would collapse into informality.¹⁹

The benchmark assumption of no commitment implies that collective choices are not possible and that the central agent has total freedom to set the fee once she takes up her role. In that case, she will behave as a revenue maximizing monopolist. However, we will also contemplate the possibility of the fee a being chosen collectively and that this choice may be binding. In this case, agents will set a fee that maximizes total welfare behind the veil of ignorance, that is, before the outcome of the lottery is realized.²⁰ Table 1 below summarizes the possible combinations of assumptions.

¹⁸It is obvious that $(\omega - a)x(a) \geq \omega \tau x(a)$ for any a not greater than the upper bound on a , which is $a(N) = \omega(1 - \tau)$.

¹⁹Note that because all agents are alike and either all or none participate in the lottery *ex ante*, we do not have to worry about the case of an agent not participating in the lottery but willing to become formal *ex post*.

²⁰Admittedly there may be other processes. The ones considered here are polar cases.

		Lottery choice is binding	
		No	Yes
Center can commit to a fee a	No	1. Agents' only commit to participate in the lottery ex ante. The center may renege ex post and is free to set a .	2. Agents commit ex ante to participate in the lottery and cannot renege ex post if chosen as the center.
	Yes	3. Agents commit ex ante to participate in the lottery. If chosen as the center, they may renege, but have no freedom to set a .	4. Agents commit ex ante to participate in the lottery and not to renege ex post. The center has no freedom to set a ex post.

Table 1.

Next we explore these different scenarios, starting with the natural benchmark, the “no commitment” case.

4.1 No commitment

Under the “no commitment” or fully decentralized procedure, the fee is freely set by the central agent. Hence, in addition to the ex-ante participation constraint, the ex-post one must be imposed. We know from Section 3 that the maximum fee that the institution can charge is $a(N)$. Therefore, agents will participate only if the two following conditions hold:

$$\frac{1}{N+1}(N(a(N) - c) + \omega) + \frac{N}{N+1}(\omega - a(N))x(a(N)) \geq \omega\tau x(a(N)), \quad (10)$$

$$N(a(N) - c) + \omega \geq \omega\tau x(0), \quad (11)$$

which are simply the result of rewriting the ex ante lottery participation constraint (8) and the ex post constraint of the center (9) by replacing a with $a(N)$. These two conditions are necessary for the institution to arise. Note that when $a = a(N)$, trading agents are indifferent between formality and informality. Therefore, (10) can be rewritten as:

$$N(a(N) - c) + \omega \geq \omega\tau x(a(N)), \quad (12)$$

from which it is evident that (10) is a necessary and sufficient condition for (11) to be satisfied. If it does not hold, the economy will remain in a state of informality.²¹

²¹Of course, this is only true for $a = a(N)$ and does not need to be verified for lower values of the fee.

Finally, we need to establish which fee will be set by the institution in equilibrium.

Proposition 4 *If condition (12) holds, there exists a SPE of the fully decentralized procedure of institution formation that implements full formality under the fee $a(N)$.*

There are two possible sources of inefficiency in this scenario. On the one hand, full formality is not implemented when (12) does not hold, despite the fact that it may still be efficiency enhancing. This is the case when parameters are such that the level of individual welfare (assuming equal weights) obtained under formality

$$W^F(a(N))(a(N)) = \frac{1}{N+1}(N(a(N) - c) + \omega) + \frac{N}{N+1}(\omega - a(N))x(a(N)),$$

dominates the level of welfare under full informality but is not high enough to induce ex ante participation in the lottery.

Corollary 1 (Non-emergence of efficient institutions) *Under the fully decentralized procedure, a potentially welfare enhancing institution does not arise if and only if the level of individual welfare under formality $W^F(a(N))$ is such that*

$$\omega\tau x(0) \leq W^F(a(N))(a(N)) \leq \omega\tau x(a(N)). \quad (13)$$

This inefficiency occurs in economies of intermediate size and when the status-quo level of trust τ is sufficiently high.

The proof can be found in the appendix. The lower bound in (13) determines when formality is more efficient than informality, whereas the upper bound establishes when formality is implementable. Between these bounds, the institution is welfare enhancing but it does not emerge.

In line with Dixit (2003b and 2004) and Greif (2006), Corollary 1 shows that the first type of inefficiency is more likely to occur in economies of intermediate size and with limited coordination problems (high τ). In the first place, it occurs if the size of the population is not small enough for informality to be superior, but not big enough for the institution to arise. The reason why N has to be large enough for the institution to arise comes from the fact that the center's expected revenue is increasing in N , so there is a minimum population size above which the prospect of becoming the center gives agents enough incentive to participate in the lottery. As in Bates et al. (2002), the

agent endowed with the institutional role reneges if she is not able to extract enough resources.

On the other hand, the range of parameters for which a welfare enhancing institution does not arise expands as τ increases. High trust makes the outside option of informality more attractive and undermines the dominant position of the revenue-maximizing institution. We should then observe the emergence of formal institutions in societies plagued with coordination problems and low levels of informal trust, while informal exchanges are more likely to prevail in societies with relatively high level of trust. When inefficiencies are not too severe, bearing the cost involved in solving them is not incentive compatible at the individual level, despite being socially efficient. In conclusion, even if the central agent is able to maximize revenue when setting a , the emergence of a desirable institution is not ensured.

This corollary would need some qualification if we relaxed two assumptions: First, if the central agent does not specialize completely in the institutional task and retains some ability to interact, her incentives to act as such would increase, hence reducing the range of inefficiency. Still, as long as she cannot enjoy the full benefit from interacting with others, inefficiency will persist. On the other hand, if the payoff of informal agents does not depend on the efficiency of the institution, inefficiency vanishes. But as soon as we have some degree of free-riding the result holds again.

Even if full formality is implemented, the fee set by the central agent may be too high and the first best may not be attained. A necessary condition for this second type of inefficiency to occur is a low enough degree of trust in bilateral interactions, i.e. $\tau < \tau^*$, that implies $a^F = a^* < a(N)$.

Corollary 2 (Implementable first best) *When condition (12) holds, the first best fee a^F can be implemented in a SPE of the fully decentralized procedure of institution formation for high enough levels of status quo-trust, i.e. $\tau \geq \tau^*$.*

The intuition for this result is easy to grasp. When welfare is increasing over the range of fees compatible with formality or, in other words, when the level of status-quo trust τ is sufficiently high, the planner would like to set the highest fee possible (i.e., $a^F = a(N)$). In that case, the center's incentives are aligned with social welfare and the first best can be attained by means of the decentralized procedure. Otherwise, the resulting fee will be excessive from the social point of view.

In conjunction, these two corollaries imply that different societies are characterized by different inefficiencies. Institutions are more likely to emerge

in societies with low levels of trust, but if they do, they will tend to be too extractive relative to the socially optimal outcome, as in Gambetta's (1993) account of the Sicilian mafia.²² On the other hand, societies with higher levels of trust may find it difficult to generate formal coordinating institutions, but if they succeed, these are more likely to be efficient.

4.2 Partial commitment

While the no commitment case appears to be the natural benchmark of our economy, it is useful to consider how the outcome of the procedure of institution formation varies when some degree of commitment is introduced along each of the two dimensions considered above: Individual commitment and a binding collective choice of the fee. Of course, that raises the question of how such a commitment is secured and enforced. We have some sort of a chicken-and-egg problem here: We started in an institutionless world, where there was a basic problem of enforcing coordination in bilateral relations. The possibility of commitment in the present case would however indicate the existence of perhaps a mechanism capable of enforcing it. After showing briefly how commitment may improve efficiency in the institution formation process under each of the possible combinations of assumptions considered in Table 1 above, we discuss how it may be enforced: In Section 5, we analyze in more detail two mechanisms that may endogenously support some degree of collective commitment despite full decentralization.

As mentioned, introducing commitment at the individual level amounts to assume that agents do not renege *ex post*, whatever the outcome of the lottery. Therefore, only agents' *ex-ante* participation constraint (8) needs to be satisfied (Case 2). On the other hand, at the collective level, commitment arises if the fee a is fixed by all participating agents before the actual running of the lottery and this choice is binding (Case 3). Finally, combining the two yields the possibility of full commitment (Case 4).

Case 2. First, assume that agents are able to commit to set up the institution if chosen to run it, so the *ex-post* participation constraint (11) is dropped, but the center retains total freedom to set the fee. Therefore, only condition (10) must hold. Since we know from Case 1 that condition (10) is stronger than (11), it is obvious that this does not introduce any change with respect to the benchmark no-commitment case. This shows that a stronger individual

²²He also introduces the twist that the institution may actually inject distrust to increase its own profit.

commitment is only useful if accompanied by some degree of collective commitment on the choice of the fee (see Case 4 below).

Case 3. Consider now the case in which a binding choice of the fee a is made by agents in advance to the lottery, but they cannot commit *ex-ante* not to renege *ex-post* in case they are chosen to run the institution. Then, society will choose a fee that maximizes social welfare subject to the ex post participation constraint, that is, a fee high enough to compensate the central agent. Hence it must be at least greater than \underline{a} . Still, the fee chosen has to be compatible with full formality (hence not above $a(N)$). Once this holds, society will implement a fee as close as possible to the first best.

Proposition 5 *The collective choice of the fee implements full formality if and only if the minimum fee that satisfies the ex-post participation constraint is smaller than the maximum fee compatible with formality, i.e. $\underline{a} \leq a(N)$. In that case, the fee set is $a = \max\{\underline{a}, a^F\}$ and the first best is achieved if and only if $\underline{a} \leq a^F$.*

First, it is important to note that the collective choice of the fee makes the implementation of the institution no easier than under the fully decentralized procedure, as it still requires $\underline{a} \leq a(N)$. However, this type of commitment makes the institution more efficient when implemented, because the first best is now more likely to be attained. On the other hand, even if that cannot be the case, i.e. $a(N) > \underline{a} \geq a^*$, there is an improvement with respect to the same case under the fully decentralized procedure, since the fee chosen is \underline{a} instead of $a(N)$, and thus closer to the social optimum.

Case 4. Finally, consider the case where there is no *ex-post* participation constraint (strong individual commitment) and agents agree that they should implement the first best.²³ It is easy to see that then the efficient outcome is always implemented.

Proposition 6 *When both individual and collective commitment are possible, full formality is implemented if and only if informality does not maximize welfare, i.e. $x(0) \leq x$. Moreover, the first-best is always attained.*

²³While in the no commitment case discussed in the previous section the *ex-post* participation constraint was irrelevant as it was implied by the *ex ante* one, this may of course not be the case when $a < a(N)$.

The intuition is straightforward: When $x(0) \leq \underline{x}$, the first best fee a^F is high enough to ensure that the *ex-ante* participation constraint (8) is satisfied. Therefore, individual incentives do not stand in the way of efficiency in this case and formality is implemented whenever it is efficient. This result is similar to the one in Acemoglu (2003), where full commitment of the ruler and citizens ensures that a political Coase theorem holds.

To summarize, when considering the decentralized institution formation process, the inability to constrain the center to chose a specific level of fee (lack of collective commitment) is a strong reason for the occurrence of inefficiencies, and one that cannot be alleviated by introducing individual commitment (Case 2). As this limit is relaxed, potential inefficiencies are reduced, as shown by Case 3. Finally, when the ability to set fees *ex-ante* is combined with individual commitment (Case 4), the first best is always implementable.

The next Section discusses two decentralized mechanisms that can enforce some degree of commitment: The use of threshold strategies in the benchmark version of our game and the possibility of coalitional secession.

5 Endogenous commitment

Threshold strategies

In the discussion above, we left aside the full informality equilibrium. Recall from Proposition 2 that full informality can arise in equilibrium of any fee in the interval $[0, a(N)]$. Agents can use this multiplicity of equilibria to tame the central agent and make him reduce the fee. To show this consider the following threshold strategy to be used by agents in our one-shot game:

$$\mathcal{F} = \begin{cases} 1 & \text{if } a \leq a' \\ 0 & \text{otherwise,} \end{cases} \quad (14)$$

where $a' \leq a(N)$. These strategies constitute an equilibrium of the subgame that starts after the central agents announces the fee. Then, if all agents use them to decide whether to become formal or not, the central agent's best response under no commitment is to choose a' . This threat enlarges the set of possible fees that can be supported in equilibrium and opens the door to a welfare improvement. Still, the *ex-post* constraint must hold, so this profile can only implement fees greater than \underline{a} , the minimum fee that satisfies the *ex-post* participation constraint.

Proposition 7 *If condition (8) holds for $a^* \geq \underline{a}$, there exists a SPE of the fully decentralized procedure that implements the first-best fee.*

Notice that the use of this type of strategies leads to a scenario very similar to the one in Case 3 because they act as a collective commitment device capable of enforcing a specific level of the fee, i.e. the threshold a' , that may be closer to the first-best fee. For that threshold to be implemented it must be compatible with the center acting as such, i.e. $a' \geq \underline{a}$, and it must ensure that agents want to participate in the process, i.e. it must satisfy condition (8). Moreover, even if the first best cannot be attained (because $a(N) > \underline{a} \geq a^*$), threshold strategies can help to reduce the inefficiency associated with a too extractive central agent. However, it is not clear how in a state of nature that we define as completely noncooperative, agents can coordinate in the use of these strategies, which seem to involve some degree of multilateral agreement.

Secession

Let us now consider the possibility that a coalition of agents secedes from society to run their own institution. Since our starting point is a state of nature where no commitment is possible, the concept of secession-proofness has a clear importance. An institution can hardly be called self-enforcing if a subgroup of agents can improve its situation by withdrawing from it.

Specifically, our analysis of secession will concentrate on the secession-proofness of the decentralized procedure of institution formation, assuming that it will be employed both by the whole population and any subgroup intending to withdraw. Then, we analyze when the threat of secession can prevent the emergence of a single institution and its effect on efficiency.

Let us first state our definition of blocking:

Definition 1 *Denote by a_N the fee set by the institution. A coalition formed by S interacting agents is a blocking coalition if and only if*

$$(\omega - a_N)x(a_N) < \frac{1}{S}(S(a(N) - c) + \omega) + \frac{S-1}{S}(\omega - a(N))x(a(N)). \quad (15)$$

Our concept of blocking implies that no group of agents should prefer (in expectation) to withdraw from society and apply among them the fully decentralized procedure of institution formation. This is a relatively strong requirement.²⁴ Note that when a coalition contemplates the possibility of secession, it recognizes that the fee that will be set in the hypothetical new

²⁴Alternatively, we could have imposed a weaker criterion, as in Howe and Roemer (1981), in which a coalition is blocking whenever it can *guarantee* a higher payoff to its members

institution must be itself self-enforcing. We have in this case picked $a(N)$ the equilibrium fee we have at length considered in the previous sections.²⁵

Definition 2 *A fee a_N is said to be secession-proof if it does not spawn any blocking coalition.*

Secession-proof fees are natural focal points in the process of institution formation: They are said to be in the *core* of that particular procedure of institution formation.²⁶

Given that we are analyzing the case of no commitment, we assume that the central agent will set the maximum possible secession-proof fee. Secession thus imposes new and natural constraints on the fee that the institutional agent can charge. Notice that, if full formality is not implementable when secession is not an option, this will continue to be the case when secession is possible; since the revenue of the central agent cannot increase, secession thus cannot help potentially welfare enhancing institutions to emerge.

The first question that arises is whether the set of secession-proof fees is empty or not. It is easy to check that the payoff of a coalition contemplating the possibility of withdrawing is increasing in its size S . Therefore, for a fee to be secession-proof it is enough to satisfy condition (15) for $S = N$.

On the other hand, the fee that maximizes agents' welfare solves

$$\begin{aligned} \max_a \quad & N(\omega - a)x(a) \\ \text{s.t.} \quad & a \leq a(N). \end{aligned}$$

The above program yields an interior solution a^{**} characterized by the first order condition

$$\omega - a^{**} = \frac{x(a^{**})}{x_a(a^{**})}. \quad (16)$$

Therefore it is clear that $a^{**} < a^*$. Again, there exist a threshold τ^{**} such that the solution to this problem is interior whenever $\tau \geq \tau^{**}$. It is straightforward as well to show that $\tau^* < \tau^{**}$. Hence, the level of the fee that maximizes the welfare of the set of interacting agents is either a^{**} or $a(N)$. Let us assume, for the sake of exposition, that $\min\{a^{**}, a(N)\} > \underline{a}$.

²⁵Note that for all S , $a(N) = a(S) = \omega(1 - \tau)$, so we stick to the current notation for simplicity.

²⁶As any core-related concept, our definition of blocking only takes into account one-step secessions. We do not consider the possibility of further blocking once a new society is formed. The set of secession-proof fees defined here is thus minimal in this sense.

The set of secession-proof fees is thus non-empty if and only if

$$\begin{aligned} & \frac{1}{N}((N-1)(a(N)-c)+\omega) + \frac{N-1}{N}\tau\omega x(a(N)) \\ & \leq (\omega - \min\{a^{**}, a(N)\})x(\min\{a^{**}, a(N)\}). \end{aligned} \quad (17)$$

The left hand side of this expression is the expected benefit from forming a coalition of N agents and seceding from the institution. The right hand side is the benefit from staying under the current institution when the central agent sets the fee that maximizing the payoff of the rest of agent. Thus, if this condition is not met, a single institution cannot be stable and we should expect multiple ones to emerge. The next proposition summarizes the conditions, in terms of the population size N and the level of status-quo trust τ , under which secession may occur.

Proposition 8 *The set of secession-proof fees is non-empty if and only if N is below a certain threshold \tilde{N} . Furthermore, \tilde{N} attains a minimum at $\tau = \tau^*(< \tau^{**})$.*

The main reason for blocking in this model is thus the prospect of becoming the center in the new mini society. When the size of the population is sufficiently big, the center obtains an extremely high payoff and this creates strong incentives to withdraw. As a matter of fact $a(N)$ stops being secession-proof when condition

$$(N-1)(a(N)-c)+\omega > \tau\omega x(a(N)),$$

holds, implying that the central agent of the new institution can obtain a higher payoff than the rest of agents.

Figure 4 depicts the regions characterized by these thresholds in the parameter space.

When the level of status-quo trust is sufficiently small (i.e. $\tau < \tau^{**}$) and the population size is intermediate, it may still be possible for the institution to avoid secession by charging a fee below $a(N)$. In that case, secession can help to alleviate the inefficiency produced by a too high fee compared to the case where secession is not possible. But outside this region of parameters, secession is a real threat that renders impossible the emergence of one institution comprising all agents in society.

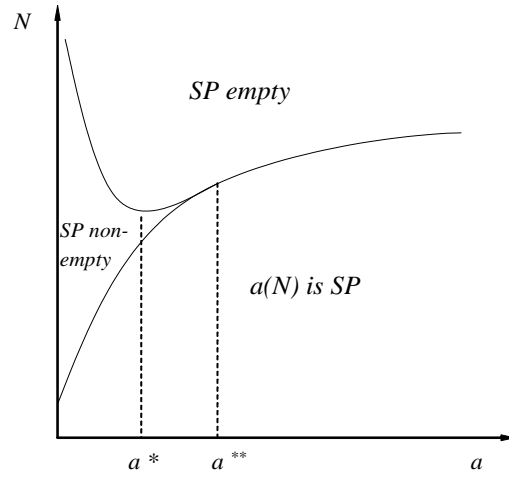


Figure 4: The set of secession-proof fees.

The natural question that now arises is whether the impossibility of a single institution matters from an efficiency perspective. The answer of course depends on the particular rules of secession and coalition formation to be considered. Here we will assume that whatever this process is, any division of the population into several smaller societies is stable only if all groups can set a secession-proof fee.

Formally, a *coalition structure* is a division of the population into a collection $C = \{C_1, \dots, C_K\}$ of disjoint coalitions of generic size $S_k \geq 3$. It is straightforward to extend our previous definition of secession-proof fees to the case of subgroups: We will say that a coalition structure C is secession-proof if all coalitions in it set a (possibly different) fee that does not spawn a blocking coalition within them. Here, we will concentrate on the case of $\tau \geq \tau^{**}$ for simplicity, meaning that in any secession-proof coalition structure all groups will set $a(N)$ since it is the unique secession-proof fee.

Next we show that if one considers secession-proof coalition structures as the natural outcome of any process of coalition formation (or secession), the impossibility of a single institution is negative from a social point of view.

Proposition 9 *When $\tau \geq \tau^{**}$, the total sum of payoffs under the single institution is at least as big as under any secession-proof coalition structure.*

As mentioned before, the incentives to secede come from agents' prospect of becoming the center of the new mini-society. Recall that when the fee is $a(N)$, it is only the central agent who extracts positive rents. However, this

is socially wasteful because it leads to an unnecessary proliferation of institutions. Obviously, this conclusion makes abstraction from the possibility that the coordination job of the center in smaller groups may entail lower transaction costs, i.e. lower c in our model. If that is the case, a trade-off arises and the above conclusion may require qualification.

6 Discussion and conclusion

The contribution of this paper is to focus on the process through which institutions aimed at enforcing cooperation may actually emerge when such coordination device did not previously exist. More specifically, our aim is to determine whether these mechanisms arise whenever they are potentially welfare enhancing, and whether they are as efficient as they could possibly be when they do arise.

We have built a model in which economic interactions take the form of a modified prisoner's dilemma game. In a hypothetical state of nature, agents are randomly matched to play this game without any interference and hence some degree of inefficiency ensues. We have assumed that agents can delegate the task of enforcing cooperation to one of them (the institution) in exchange for a proper compensation. Examples of this type of mechanisms can be found in Economics, Sociology and Law.

In a world of no commitment, the main motivation to participate in the process of institution formation is the potential rent associated with being a revenue maximizing center. In this context, the model yields clear answers to both questions above. First, there exists a region in the parameters space in which a potentially welfare enhancing institution does not arise. This is because individual and social incentives are not aligned and each individual fails to internalize the cost that he imposes on others by opting out of the potential institutional arrangement. Such an inefficiency is more likely for societies of intermediate size, an insight reminiscent of Dixit's (2003b) "intermediate-size worlds fare worst" result. Groups that are too small are optimally left to the informal type of interaction. Although this is not made explicit in our model, an additional intuitive reason for this may be that N and the payoff from informal interactions y (and therefore the endogenous measure of trust τ) are inversely related. On the other hand, as the number of individuals grows, the rent associated with being in charge of running the coordinating institution becomes large enough to ensure that it will emerge.

Moreover, a welfare enhancing institution may fail to arise if the gap between the payoff from non cooperation and cooperation is not very large,

that is, if what we called trust in the state of nature is high enough. Because the outside option is not very bad, agents are more reluctant to engage in the costly process of institution creation. This intuitive negative correlation between the likelihood of the emergence of formal institutions and the level of trust sheds light on one of the fundamental identification problems that arise in the empirical literature on social capital (see Durlauf and Fafchamps, 2006). Indeed, it seems to be the case that when formal institutions are weak, social capital (understood for example as trust in our model) substitutes for them. When formal institutions grow stronger, a process that occurs along the path of development, some form of social capital may be destroyed or become less important (see Routledge and von Amsberg, 2003, for theoretical examples of such effects). We may therefore observe a negative correlation between measures of trust and social or economic outcomes, but rather than reflecting some causal link, it is the result of a fundamental endogenous link between social capital and formal institutions of the type described in our model.

Second, our model makes a step towards understanding the observed heterogeneity of institutions. Indeed, even when the institution emerges, it may do so at various levels of efficiency, and in particular it may be suboptimal, in the sense that it will charge a fee that is above the welfare maximizing level. This is due to the absence of a collective commitment device to set the institutional fee in advance, which allows the chosen center to adopt a revenue maximizing strategy.

However, contrary to the previous one, this type of inefficiency is more likely to happen for low levels of trust, i.e. when the gap between non cooperative and cooperative payoffs is large. So different societies face different potential problems. When trust is low, a welfare enhancing institution is likely to arise but will probably be too extractive in nature. In a sense, this is the price to pay for coordination to be enforced in a context in which the loss from non-cooperation is large. On the other hand, when trust is high, an institution may not arise, but if it does, it is more likely to be efficient. Indeed, because the gains from formal coordination are relatively low in that case, an institution that would be too extractive is unlikely to be individually incentive compatible in the first place.

We then show that the two types of inefficiencies stem from the lack of individual and collective commitment. However, there is a fundamental asymmetry here, in the sense that individual commitment to run the institution would not change the results above unless it is accompanied by collective commitment on the fee that will be charged ex post. On the other hand, collective commitment goes some way towards solving excessive rent extraction and, if

accompanied by individual commitment, it does restore the first best in the sense that a political Coase theorem can hold (Acemoglu, 2003).

The question of course is how commitment may arise endogenously in a world in which no coordination device or authority exist ex-ante. We show that the threat of secession by subgroups of agents may generate such collective commitment, at least when the level of trust is low enough and the number of agents not too large. On the other hand, as this number becomes large enough, secession becomes unavoidable, resulting in a multi-institution world. This is for example the case in the work of Bueno de Mesquita et al. (2003) where they show that increasing demands from the ruler may trigger defection from a subset of agents, as in the case of Vichy France.

In the basic version of our model, secession always reduces welfare compared to a unique central institution. However, transaction cost considerations may introduce a trade-off here, if for example coordination in smaller groups is characterized by lower such costs. Endogenizing these transaction costs is an interesting area for future research and would make it possible to better understand situations characterized by multiple institutional spheres.

The issue of collective commitment to limit the level of fee extraction opens to the broader question of how a constitution can place checks on abuses by the ruler in more complex settings. While this is beyond the scope of this paper, the work of Roger Myerson is relevant here, especially the idea that agents under this rule must be rewarded to ensure stability of power (e.g., Myerson, 2006a and 2006b). Although we do not explicitly model the agency issues considered there, Myerson's point in terms of the ruler having to commit to large enough future rewards resembles our result on the threat of secession putting limits to the ruler's extractive power.

Some further remarks regarding our assumptions are in order. One possible modification of our model is to relax the anonymity of the matching technology. If formal agents could verify the identity of their partner right after being matched nothing would change. The payoff from interacting with an informal agent is by assumption always greater than the autarchic payoff, so it never pays off to walk away from a match. On the other hand, one could assume that formal agents have a probability of matching with other formal agents higher than the one implied by the anonymous random matching process. That would increase the value of formality and hence the fee that the center can charge. However, apart from the non-generic case in which formal agents are matched to each other with probability one, the configuration of equilibria would not change.

A repeated version of the game in which the central agent stays in the role once chosen would not improve upon the one-shot version. In this context,

agents' trigger strategies would be identical to the threshold strategies we have already discussed in Section 5. Moreover, the central agent could charge fees above $a(N)$ by threatening not to perform his task. Things would be different though if the central agent could be replaced. In this case, the parameter set in which an efficient institution arises is likely to expand. This is an open question that we leave for further research.

Finally, we have assumed identical agents because we were mainly interested in the factors that may hinder or foster the emergence of institutions. It is clear, however, that individual heterogeneity represents an interesting avenue for further research and in the future we intend to explore the impact of endowment inequality on the results of the present paper. This may have interesting implications, in particular in the field of development economics.

7 Appendix

Proof of Proposition 1. Since $a(K)$ is increasing in K , only corner configurations can prevail, i.e. no intermediate number of formal agents $0 < K < N$ can be supported as an equilibrium of this stage game. Suppose that $a \leq a(K)$ so no formal agents wants to deviate. Then, since we also have $a < a(K+1)$, informal agents would deviate and become formal, leading to full formality. Similarly, if $a \geq a(K+1)$, which is the necessary condition to sustain $N - K$ informal agents, formal agents would have an incentive to defect to informality, leading to an equilibrium with only informal agents. ■

Proof of Proposition 3. The condition (7) comes from just comparing the welfare under full formality with expression (6). Taking the derivative of the right hand side with respect to N yields

$$\frac{\partial x}{\partial N} = \frac{1}{(N+1)^2} \left(\frac{(\omega - a^F) x(a^F) + a^F - c}{\tau \omega} - \frac{1}{\tau} \right),$$

where we make use of the fact that, regardless of whether the solution is interior or not, a^F does not depend on N . It can be shown that $\frac{\partial x}{\partial N} > 0$. Similarly,

$$\begin{aligned} \frac{\partial x}{\partial \omega} &= \frac{N}{N+1} \frac{a^F(x(a^F) - 1) + c}{\tau \omega^2} \\ &\quad + \frac{N}{N+1} \frac{1}{\tau \omega} \frac{\partial a^F}{\partial \omega} (-x(a^F) + (\omega - a^F) x_a(a^F) + 1). \end{aligned}$$

Note first that the expression in brackets in the second term is the FOC of the planner's problem and hence it is nonnegative. Second, if $a^F = a(N)$,

$\frac{\partial a^F}{\partial \omega} = 1 - \tau > 0$ and then it is clear that the lower bound \underline{x} is increasing in ω . On the other hand, when $a^F = a^*$ the bracketed term is equal to zero since the FOC of the planner's problem is binding. Finally,

$$\begin{aligned} \frac{\partial x}{\partial \tau} = & \frac{1}{N+1} \left(-\frac{1}{\tau^2} - N \frac{(\omega - a^F) x(a^F) + a^F - c}{\tau^2 \omega} \right. \\ & \left. + \frac{N}{N+1} \frac{1}{\tau \omega} \frac{\partial a^F}{\partial \tau} (-x(a^F) + (\omega - a^F) x_a(a^F) + 1) \right). \end{aligned}$$

Again, when $a^F = a(N)$ then $\frac{\partial a^F}{\partial \tau} = -\omega > 0$ and \underline{x} is decreasing in τ , and when $a^F = a^*$ the second term is equal to zero. ■

Proof of Corollary 1. The comparative statics on N can be derived by noting that $W^F(a(N))$ is increasing in N , while the upper and lower limits do not depend on N (since $a(N) = \omega(1 - \tau)$). Rewriting $W^F(a(N)) = \frac{N}{N+1} [a(N) - c] + (\omega - a(N)) x(a(N)) + \frac{\omega}{N+1}$, the derivative with respect to N is given by

$$\frac{\partial W^F(a(N))}{\partial N} = \frac{\omega (\tau x(a(N)) - \tau) - c}{(N+1)^2},$$

which is positive since by assumption $\omega(1 - \tau) > c$ and $\tau x(a(N)) > 1$.

On the other hand, the effects of the level of status-quo trust τ can be estimated in the following way. Differentiating

$$W^F(a(N)) = \frac{N}{N+1} [\omega - c + \tau \omega [x(\omega(1 - \tau)) - 1]] + \frac{\omega}{N+1},$$

with respect to τ , we get that

$$\frac{\partial W^F(a(N))}{\partial \tau} = \frac{N}{N+1} \omega [x(\omega(1 - \tau)) - 1 - \tau \omega x'(\omega(1 - \tau))],$$

while the derivative of the upper bound is given by:

$$\frac{\partial [\omega \tau x(a(N))]}{\partial \tau} = \omega [x(\omega(1 - \tau)) - \tau \omega x'(\omega(1 - \tau))].$$

Since $\frac{\partial [\omega \tau x(a(N))]}{\partial \tau} > \frac{\partial W^F(a(N))}{\partial \tau}$, and $W^F(a(N)) > \omega \tau x(a(N))$ for τ close to 0 (the right hand side then tends to 0), we deduce that there is a threshold value $\underline{\alpha}$ such that formality is only implemented through the decentralized procedure if $\tau < \underline{\alpha}$. Note that depending on the value of the parameters, it might be the case that $\underline{\alpha} > 1$, so no inefficiency arises. ■

Proof of Proposition 8. When $a(N) > a^{**}$, expression (17) implicitly defines a threshold on the population size denoted by $N_0(\tau, \omega)$ such that a^{**} is secession-proof whenever $N \leq N_0(\tau, \omega)$. Similarly, when $a(N) < a^{**}$ the threshold

$$N_1(\tau, \omega) \equiv \frac{\tau\omega x(a(N)) - \omega}{a(N) - c} + 1,$$

can be defined as the maximum population size that is compatible with $a(N)$ being secession-proof. Hence, denote by

$$\tilde{N} \leq \begin{cases} N_0(\tau, \omega) & \text{if } \tau \leq \tau^{**} \\ N_1(\tau, \omega) & \text{otherwise} \end{cases}.$$

Now recall from our discussion in Section 3 that there exists a value of the status quo trust denoted by τ^{**} such that $a^{**} \geq a(N)$ whenever $\tau \geq \tau^{**}$. In that case, $\min\{a^{**}, a(N)\} = a^{**}$ and the threshold $N_0(\tau, \omega)$ applies. By the Implicit Function Theorem,

$$\frac{\partial N_0(\tau, \omega)}{\partial \tau} = N(N-1)\omega \frac{1 - x(a(N)) + \tau\omega x_a(a(N))}{\tau(\omega x(a(N)) - \omega - c)}.$$

Note that the denominator is the FOC of the utilitarian planner problem. We know that when $\tau < \tau^*$ then $a(N) < a^*$, and the numerator is negative (positive otherwise).

Similarly, for $\tau > \tau^{**}$, $N_1(\tau, \omega)$ becomes the relevant threshold and

$$\frac{\partial N_1(\tau, \omega)}{\partial \tau} = \omega \frac{x(a(N))(\omega - c) - \tau\omega x_a(a(N))(a(N) - c) - \omega}{(a(N) - c)^2}.$$

Since in this case, $a(N) < a^{**}$, then $x(a(N)) > \tau\omega x_a(a(N))$ so the denominator has a positive sign. Note as well, that this derivative evaluated at $\tau = 0$ is positive, and that the denominator is decreasing in τ . Hence, $N_1(\tau, \omega)$ is everywhere increasing in τ . ■

Proof of Proposition 9. When C is secession-proof the total sum of payoffs is simply

$$\begin{aligned} W_C^F &= \sum_{k=1}^K [(S_k - 1)(a(N) - c) + \omega + (S_k - 1)\tau\omega x(a(N))] \\ &= (N + 1 - K)(a(N) - c + \tau\omega x(a(N))) + K\omega. \end{aligned}$$

This expression is clearly decreasing in K , the number of coalitions in C . Therefore, the total sum of payoffs under any secession-proof coalition structure can never be greater than under the single institution (they are equal if the single institution is secession-proof itself). ■

References

- [1] Acemoglu, D. 2003. Why Not a Political Coase Theorem? Social Conflict, Commitment and Politics. *Journal of Comparative Economics*, **31**, 620-652.
- [2] Acemoglu, D. and J. Robinson. 2005. *Economic Origins of Dictatorship and Democracy*. Cambridge University Press.
- [3] Acemoglu, D., J. Robinson and T. Verdier. 2004. Kleptocracy and Divide-and-Rule: A Model of Personal Rule. *Journal of the European Economic Association Papers and Proceedings*, **2**, 162-192.
- [4] Ahlerup, P., Olsson, O. and Yanagizawa, D. 2009. Social Capital vs. Institutions in the Growth Process. *European Journal of Political Economy*, **25**, 1-14.
- [5] Aoki M., K. Murdock, and M. Okuno-Fujiwara. 1997. Beyond the East Asian Miracle: Introducing the Market Enhancing View. In M. Aoki, H. Kim and M. Okuno-Fujiwara, Eds., *The Role of Government in East Asian Economic Development: Comparative Institutional Analysis*. Oxford University Press: Oxford.
- [6] Attali, J., 2003. *L'Homme nomade*. Fayard.
- [7] Azuma, Y. and H. Grossman. 2002. A Theory of the Informal Sector. NBER working paper 8823.
- [8] Bagwell, K. and Staiger R. W. 1990. A Theory of Manage Trade. *American Economic Review*, **80(4)**, 779-795.
- [9] Baker, M. J. 2003. An Equilibrium Conflict Model of Land Tenure in Hunter-Gatherer Societies. *Journal of Political Economy*, University of Chicago Press, **111(1)**, 124-173.
- [10] Baker, M. Bulte, E. and Weisdorf, J. 2010. The Origins of Governments: From Amorphy to Anarchy and Hierarchy. Forthcoming, *Journal of Institutional Economics*.
- [11] Bardhan, P.K. 2005. *Scarcity, conflicts and cooperation: essays in the political and institutional economics of development*. MIT Press: Massachusetts.

- [12] Basu, K. 2000. *Prelude to Political Economy*. Oxford University Press: Oxford.
- [13] Bates, R., A. Greif, and S. Singh. 2002. Organizing Violence. *Journal of Conflict Resolution*, **46**, 599-628.
- [14] Bernstein, L. 2001. Private Commercial Law in the Cotton Industry: Creating Cooperation through Rules, Norms, and Institutions. *Michigan Law Review*, **99**, 1724-1788.
- [15] Bernstein, L. 1992. Opting out of the Legal System: Extralegal Contractual Relations in the Diamond Industry. *Journal of Legal Studies*, **21**, 115-157.
- [16] Bueno de Mesquita, B., Smith, A., Siverson R. M., Morrow, J. D. 2003. *The Logic of Political Survival*. MIT Press: Massachusetts.
- [17] Choi, J.P. and M. Thum. 2002. Corruption and the Shadow Economy. CESifo Working Paper 633.
- [18] Dixit, A. 2003a. On Modes of Economic Governance. *Econometrica*, **71(2)**, 449-481.
- [19] Dixit, A. 2003b. Trade Expansion and Contract Enforcement. *Journal of Political Economy*, **111**, 1293-1317.
- [20] Dixit, A. 2004. *Lawlessness and Economics: Alternative Modes of Governance*. Princeton University Press.
- [21] Durlauf, S. and M. Fafchamps. 2006. Social Capital. In *Handbook of Economic Growth*, P. Aghion and S. Durlauf, Eds., North Holland: Amsterdam.
- [22] Ellison, G. 1994. Cooperation in the Prisoner's Dilemma with Anonymous Random Matching. *Review of Economic Studies*, **61**, 567-588.
- [23] Ensminger, J. 1992. *Making a Market. The Institutional Transformation of an African Society*. Cambridge University Press: New York.
- [24] Fafchamps, M. 2004. *Market Institutions in Sub-Saharan Africa*. MIT Press: Massachusetts.
- [25] Gambetta, D. 1993. *The Sicilian Mafia: The Business of Protection*. Harvard University Press: Massachusetts.

- [26] Greif, A. 1993. Contract Enforceability and Economic Institutions in Early Trade: The Maghribi Traders' Coalition. *American Economic Review*, **83**, 525-48.
- [27] Greif, A. 1994. Cultural Beliefs and the Organization of Society: A Historical and Theoretical Reflection on Collectivist and Individualist Societies. *Journal of Political Economy*, **102**, 912-50.
- [28] Greif, A. 1997. Microtheory and Recent Developments in the Study of Economic Institutions through Economic History. In D.M. Kreps and K. F. Wallis, Eds., *Advances in Economic Theory (vol. 2)*. Cambridge University Press: New York.
- [29] Greif, A. 2005. Commitment, Coercion, and Markets: The Nature and Dynamics of Institutions Supporting Exchange. In C. Menard and M. M. Shirley, Eds., *Handbook for New Institutional Economics*. Kluwer Academic Publishers: Norwell MA.
- [30] Greif, A. 2006. *Institutions and the Path to the Modern Economy: Lessons from Medieval Trade*. Cambridge University Press: New York.
- [31] Grossman, H. 2002. "Make us a king": Anarchy, Predation, and the State. *European Journal of Political Economy*, **18**, 31-46.
- [32] Howe, R.E. and J.E. Roemer. 1981. Rawlsian Justice as the Core of A Game. *American Economic Review*, **71**, 880-895.
- [33] Kandori, M. 1992. Social Norms and Community Enforcement. *Review of Economic Studies*, **59**, 63-80.
- [34] Konrad, K. and W. Leininger. 2007. Self-enforcing Norms and the Efficient Non-cooperative Organization of Clans. *Public Choice*, forthcoming.
- [35] Kranton, R. 1996. Reciprocal Exchange: A Self-sustaining System. *American Economic Review*, **86**, 830-851.
- [36] Kreps, D., P. Milgrom and R. Wilson. 1982. Rational Cooperation in the Finitely Repeated Prisoners' Dilemma. *Journal of Economic Theory*, **27**, 245-252.
- [37] Li, S. 2003. The Benefits and Costs of Relation-based Governance: An Explanation of the East Asian Miracle and Crisis. *Review of International Economics*, **11**, 651-667.

- [38] Marcouiller, D. and L. Young. 1995. The Black Hole of Graft: The Predatory State and the Informal Economy. *American Economic Review*, **85**, 630-646.
- [39] McMillan, J. and C. Woodruff. 2000. Private Ordering under Dysfunctional Public Order. *Michigan Law Review*, **98**, 2421-2458.
- [40] McMillan, J. and C. Woodruff. 1999. Dispute Prevention without Courts in Vietnam. *Journal of Law, Economics & Organization*, **15**, 637-658.
- [41] Milgrom, P., D. North and B. Weingast. 1990. The Role of Institutions in the Revival of Trade: The Medieval Law Merchant, Private Judges and the Champagne Fairs. *Economics and Politics*, **1**, 1-23.
- [42] Moselle B. and B. Polak. 2001. A Model of a Predatory State. *Journal of Law, Economics, and Organization*, **17**, 1-33.
- [43] Myerson, R. B. 2006a. Fundamental Theory of Institutions: A Lecture in Honor of Leo Hurwicz. Hurwicz Lecture, North American Meetings of the Econometric Society.
- [44] Myerson, R. B. 2006b. Leadership, Trust, and Constitutions. mimeo, University of Chicago.
- [45] North, D. 1990. *Institutions, Institutional Change, and Economic Performance*. Cambridge University Press: Cambridge.
- [46] Okzaki, T. 1997. The Government-Firm Relationship in Postwar Japanese Economic Recovery: Resolving the Coordination Failure by Coordination in Industrial Rationalization. In M. Aoiki, H. Kim and M. Okuno-Fujiwara, Eds., *The Role of Government in East Asian Economic Development: Comparative Institutional Analysis*. Oxford University Press: Oxford.
- [47] Olson, M. 1965. *The logic of Collective Action*. Harvard University Press: Massachusetts.
- [48] Olson, M. 2000. *Power and Prosperity: Outgrowing Communist and Capitalist Dictatorships*. Basic Books.
- [49] Rawls, J. 1971. *A Theory of Justice*. Harvard University Press: Massachusetts.

- [50] Rotemberg J. J. and Saloner, G. 1986. A Supergame-Theoretic Model of Price Wars during Booms. *American Economic Review*, 76(3), 390-407.
- [51] Richman, B.D. 2006. How Community Institutions Create Economic Advantage: Jewish Diamond Merchants in New York. *Law and Social Inquiry*, **31**, 383-420.
- [52] Ross, I.S. 1995. *The Life of Adam Smith*. Clarendon Press: Oxford.
- [53] Routledge B. and J. von Amsberg. 2003. Social Capital and Growth. *Journal of Monetary Economics*, **50**, 167-193.
- [54] Straub, S. 2005. Informal Sector: The Credit Market Channel. *Journal of Development Economics*, **78**, 299-321
- [55] Taylor, M. 1982. *The Possibility of Cooperation*. Cambridge University Press: Cambridge.